SFA Modernization Partner
United States Department of Education
Student Financial Assistance

Data Warehousing Implementation Strategy
NSLDS Technical Architecture Analysis

Task Order 21
Deliverable 21.1.5

July 31, 2000

Contents

Introduction

This analysis is the second part of a two-part study of the National Student Loan Data System (NSLDS). While the first part – NSLDS Cost Analysis, delivered on 6/30/00 – focused on presenting NSLDS' cost components, this analysis presents NSLDS' technical architecture.

In presentating the technical architecture, we took the approach of decomposing NSLDS' logical architecture presented in the NSLDS Cost Analysis document (also shown on the next page). More specifically, the areas we studied are:

- Database architecture
- Data population architecture
- Application architecture
- End-user access architecture
- Operations architecture

We begin the discussion with a detailed background of NSLDS, followed by a section on each architecture, and conclude with an analysis of our technical architecture findings.

Scope

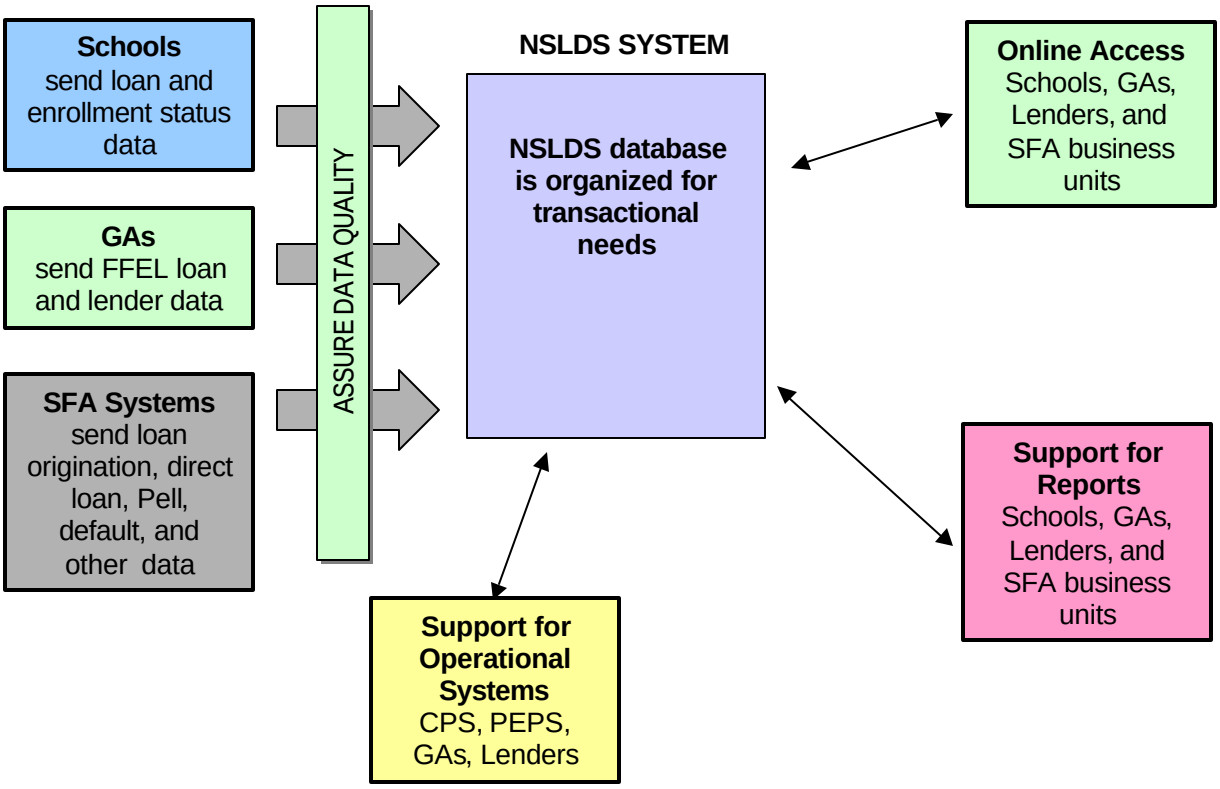
This analysis presents a summary of each of the above architectural areas, and ends with ideas for improving NSLDS and its value. More detailed information can be obtained from the original NSLDS system documentation if necessary.

This document is intended for use by SFA executives as they make decisions on modernizing systems and developing new capabilities.

Background

The National Student Loan Data System (NSLDS) is a transactional system – not just a repository – that serves several important business functions, such as checking student and school eligibility. NSLDS, initiated by congressional mandate, has been in production since 1994, and contains student aid data going back to 1965.

Shown below, NSLDS imports student aid data from about 2,900 schools, 36 guaranty agencies (GAs) representing 5,000 lender and loan servicers, and 5 internal SFA systems. The aid data is collected primarily to serve system-to-system business functions (described later in this document), but since the data is already centralized, it serves users through predefined and ad-hoc online queries and reports.



A logical architecture view of the NSLDS system

NSLDS consolidates Title IV student aid information into a single repository, and contains borrower and grant information. The repository includes the:

- William D. Ford Federal Direct Loan Program

- Federal Family Education Loan (FFEL) Program
- Federal Insured Student Loan (FISL) Program
- Federal Perkins Loan Program (including National Defense Student Loans, National Direct Student Loans, Perkins Expanded Lending and Income Contingent Loans)
- Federal Pell Grants
- Overpayments on a Federal Pell Grant, Federal Supplemental Educational Opportunity Grant or Federal Perkins Loan

NSLDS' Value Proposition

Before NSLDS came into existence, Federal student aid was delivered, administered, and collected by private lenders, loan servicers and schools. The Department of Education provided the funding and financial guarantees but had little insight into the effectiveness of the overall operation. Aid information resided on thousands of lender, servicer, guaranty agency, and school systems. Moreover, the information was of poor quality, and the only reporting on the $100 billion outstanding guaranteed loans was done once a year, based on information from guaranty agencies.

Congress directed the Department to develop a consolidated repository of student aid information specifically to reduce rapidly increasing defaults (which was approaching 25%) and fraud.

NSLDS' was conceived with three main goals:

- To improve the quality and accessibility of student aid information
- To reduce the burden of administering Title IV aid
- To minimize abuse of aid programs through more accurate tracking

Early Challenges

NSLDS' design team faced difficult challenges related to obtaining external data: there were thousands of systems, mostly with poor data quality; their systems were not built to support electronic data exchange; there were no consistent standards for identifying a loan or a student; their systems were slow and low technology; and there was a lack of interest and incentive to fund participation.

Today, NSLDS is operational and centralizes information from schools, guaranty agencies (GAs centralize lender information), and internal SFA systems. By creating a central repository, NSLDS enables the Department to more effectively manage and deliver student aid through Title IV. The next section describes NSLDS' capabilities.

Business Functions

NSLDS serves a number of business functions as a result of centralizing student aid data. These functions are delivered to other systems and end users. Users can access and view many of these functions through the internet or the TIVWAN, and can run predefined reports, data extracts, and can submit reports / extracts to be run in batch. They can also view some frequently used information online.

The following is a list of student aid functions:

- Provide pre-screening and post-screening for Title IV aid eligibility
- Calculate default rates for educational institutions, guaranty agencies, and lenders
- Monitor the reasonability of payments to guaranty agencies and lenders
- Report changes in student / borrower enrollment status
- Prepare electronic financial aid history information
- Assist guarantors, schools, servicers, and financial institutions in collecting loans
- Provide audit and program review planning
- Support research and policy development
- Conduct budget analysis and development
- Track loan transfers from one organization to another
- Track loan borrowers and overpayment debtors
- Provide information that supports Credit Reform Act of 1992 requirements
- Provide information to track the timely processing of refunds and cancellations
- Assist SFA in collecting debts
- Assess administration of guarantors, schools, financial institutions, and servicers

The most important of these functions are described in more detail later in this section. Other what is listed above, NSLDS serves as the entry point for 77% of the student aid information that is external to SFA. By putting a quality process in place, NSLDS helps to assure data quality of external systems.

The table below shows how (system-to-system, reports, online access) NSLDS delivers many of the functions that are listed above.

Mode	*Business Function*
1	Provide pre-screening and post-screening for Title IV aid eligibility
1,2	Calculate default rates for educational institutions, guaranty agencies, and lenders
2,3	Monitor the reasonability of payments to guaranty agencies and lenders
1,2	Report changes in student/borrower enrollment status
2,3	Prepare electronic financial aid history information
2,3	Assist guarantors, schools, servicers, and financial institutions in collecting loans
2,3	Provide audit and program review planning
2,3	Support research and policy development
1,2,3	Conduct budget analysis and development
1,2,3	Track loan transfers from one organization to another
2,3	Track loan borrowers and overpayment debtors
2,3	Provide information that supports Credit Reform Act of 1992 requirements
3	Provide information to track the timely processing of refunds and cancellations
1,2,3	Assist SFA in collecting debts
1,2,3	Assess administration of guarantors, schools, financial institutions, and servicers

Category (1= Support for Operational System to System, 2=Support for Reports, 3= Online Access)

The following section describes some of the business functions of NSLDS in more detail.

Student Eligibility - Pre-screening and post-screening

NSLDS summarizes all previous Title IV aid a student has received into a financial aid history that is used by school Financial Aid Administrators to determine student eligibility and to configure a financial aid package for the student. NSLDS pre-screens all applicants for Title IV aid to identify applicants who are in default on an existing Title IV loan; who owe overpayments on Pell grants, FSEOGs, or Perkins loans; or who have already borrowed the maximum amount allowed under the annual, or cumulative loan limits for each loan type. NSLDS post-screens Title IV aid applicants to identify applicants whose eligibility changes after the time of their original aid application was pre-screened. This function is mandated by legislation and has a yearly estimated savings/cost avoidance of $575 million. The financial aid history is also available to authorized users on both a batch and online (real time) basis to handle mid-semester transfer students.

School Eligibility (Default Calculation)

The Schools channel is responsible for Draft and Official Cohort Default Rates which is required by legislation to be calculated semi-annually (monthly for notional default rates) for schools participating in the Federal Direct Loan Program and Federal Family Educational Loan Program. Rates are also calculated for guaranty agencies and lenders. Schools with default rates 25% or higher for at least three consecutive years (40% for 1 year) can be disqualified from participating in some or all student financial aid programs. Since the institution of the default

calculation process, 1,180 schools have lost eligibility. NSLDS makes these rates (and loan detail) available on the web to Schools and selected SFA users.

Student Enrollment Tracking

Student Enrollment Tracking performs the collection and dissemination of Student Status Confirmation Report (SSCR) information. Loan holders use the SSCR to verify a borrower's enrollment status. This enrollment information permits loan holders to perform the critical steps of placing a borrower into repayment, initiating repayment grace periods, and extending in-school deferments. NSLDS has standardized the SSCR process by creating a single consolidated roster for schools to update and by sending loan holder organizations a consolidated enrollment status file of information about their borrowers. NSLDS maintains both the current and historical enrollment information for a student.

Student Aid Tracking

Student Aid Tracking has four separate sub-functions. The Student Account enables borrowers to see a consolidated view of their account and identify the current lender, servicer, or holder of each loan. Students have access to information about their financial aid at the "NSLDS Student Access" secure web site (www.nslds.ed.gov). Students can also call the Public Inquiry Contractor's (PIC) toll free number and have the PIC operators look up the information in NSLDS. The second, Borrower Tracking, helps qualified users locate borrowers who have defaulted on student loans by identifying other schools, guaranty agencies, and lenders previously associated with a borrower so they can be contacted for the borrower's current address. Loan Transfer Tracking, monitors transfer activity by maintaining dates of sale and names of loan holders. This information identifies likely problems with participants and helps evaluate the administration and billing by lenders and guaranty agencies in the FFEL loan program. Also, NSLDS tracks Organization Contacts by their job function and helps users quickly locate the correct contact person at a school, guaranty agency, lender, servicer, or other data provider. This contact capability is accessible by both online screens and through the NSLDS Information for Financial Aid Professionals web site (www.nsldsfap.gov).

Financial Management (GA Fees Calc, Financial Reports, External Audits)

NSLDS performs the calculations for the Financial Partners channel to determine fees paid to lenders and guaranty agencies. Automatic payment of Loan Processing and Account Maintenance fees from NSLDS data replaces part of the Lender and GA Billing process (form 1130/1189) which required time-consuming invoicing and verification. Financial Statements for CFO reporting to Congress and the public are also produced from NSLDS. Finally, trend analysis and other monitoring reports are produced to retrieve specific data from NSLDS on organizations (schools, lenders, and guaranty agencies) to identify key indicators used to schedule school and financial partner Audits and Program Reviews.

Budget Forecasting and Credit Reform Act Support

Every year, CFO develops input for the President's budget, based partly on projected loan program costs for a seven-year period. NSLDS information is used to develop reliable, sound forecasts and program estimates for the Department of Education Budget; answer budget related questions; and support necessary hypothetical analyses.

The Credit Reform Act (CRA) requires loan level tracking of all federally guaranteed loans. NSLDS tracks and reports loans by program, cohort year, and risk category. Loan data is used semi-annually by CFO to estimate government costs associated with loan programs.

Research, Policy Analysis and Performance Assessment

NSLDS provides several types of access in support of internal (CFO, OIG) and external (CBO, GAO) users performing Research, Policy Development and Assessment of the Performance of various Title IV aid delivery system participants and aid programs. Online queries range from focused queries, pertaining to a single student or guaranty agency for relatively small amounts of data, to queries requiring NSLDS to supply or summarize massive amounts of data.

Guaranty agencies, schools, and lenders are provided reports for researching and assessing their own performance in administering the FFELP aid programs. Such research can be either short- or long-term and generally aims to evaluate the effectiveness of specific organizations and program practices.

Statistically valid random samples of the entire database are also created each quarter and loaded into a separate database called the Statistical Abstract (STAB). This smaller subset of the data is based upon 1.5 million students and consists of about 5 million records that accurately reflects the universe of 40 million students and the 135+ million loans and grants issued under the various loan programs.

NSLDS also computes monthly, quarterly, and annual Aggregates, which are pre-computed totals of the data. The aggregates are for frequently requested data where it is more efficient to compute the answer once and store it, than to add it up each time a user requests the answer. About 1,500 aggregates are computed each month.

NSLDS Users

The table below shows NSLDS users by business function. The names of the organization may have changed, but many of these groups/functions exist under the new SFA organization.

SFA User Community*	Acronyms*	Prescreening/Postscreening for Eligibility	Default Rate Calculation	Monitor GA & Lender Billings	Research & Policy Development	Budget Analysis & Development	Audit and Program Review Planning	Assess FFEL Admin by GAs, Schools, Lend
Analysis & Forecasting Div.	OPE / PTAS / AFD				x	x		
Applicant Systems Division	OPE / PSS / ASD	x						
Credit Mgt. Improvement Staff	OM / CFO / CMIS					x		
Div. of Planning & Eval. Svcs.	OM / CFO / BSD / DPES				x	x		
Field Operations Service	OPE / FOS						x	
General Accounting Office	GAO							
GA & Lender Oversight Staff	OPE / GLOS		x				x	
Institutional Monitoring System	OPE / IPOS / IMD		x				x	
Institutional Participation Div.	OPE / IPOS / IPD				x			
Office of General Council	OGC				x			
Office of Inspector General	OIG				x		x	
Office of Mgt. & Budget	OMB				x	x		
Policy Development Div.	OPE / PTAS / PPD				x	x		
Postsecondary Analysis Div.	OM / CFO / PAD					x		x
Postsecond. Educ. Participant Sys.	OPE / PSS / DID / PEPS		x					
Program Budget Execution Br.	OM / CFO / FSD / PBEB							
Program Systems Service	PSS			x				
Student Receivable Division	DCS / SRD							
Training & Program Info. Div.	OPE / PTAS / TPID				x			
Quality Improvement & Operations Planning Staff	OPE / QIOPS				x			
System Coordination Branch	SCB	x						
System Planning Branch	SPB		x					
Acctng. & Finance Mgt. Section	AFMS			x				
Office of Ed. Research & Improv.—Nat'l. Ctr. for Ed. Stat.	OERI / NCES				x			

** These user group/acronym names are those before the reorganization of SFA.*

Database Architecture

NSLDS's data repository stores transaction-level detail on loans and default rates, with referential relationships with students, guaranty agency, lenders and school entities. NSLDS was designed to be an online transaction processing (OLTP) system to handle the day-to-day operations of particular business functions. The database is a classic OLTP design with the following characteristics:

- Normalized data structures (3rd normal form)
- Referential integrity for assuring data quality
- Transaction-level data
- Efficiency in data storage
- Efficiency in executing the functions it was designed to serve

Since NSLDS imports and processes an unusally high volume of data from so many external data providers, the database is designed with high-volume transactions in mind. See Appendix B (Database Volumetric data) for record counts by table.

There are mutiple databases within NSLDS which were created to serve different purposes as follows:

Active Database – This database holds loan-level information for *open* student/borrower accounts. Each of the NSLDS functional requirements and most of the functional capabilities are satisfied by this database.

Archive Database – This database was designed to hold loan-level information for *closed* student/borrower accounts. However, the database has never been populated with data, and is currently empty. SFA wants all of NSLDS' information to be in the Active database.

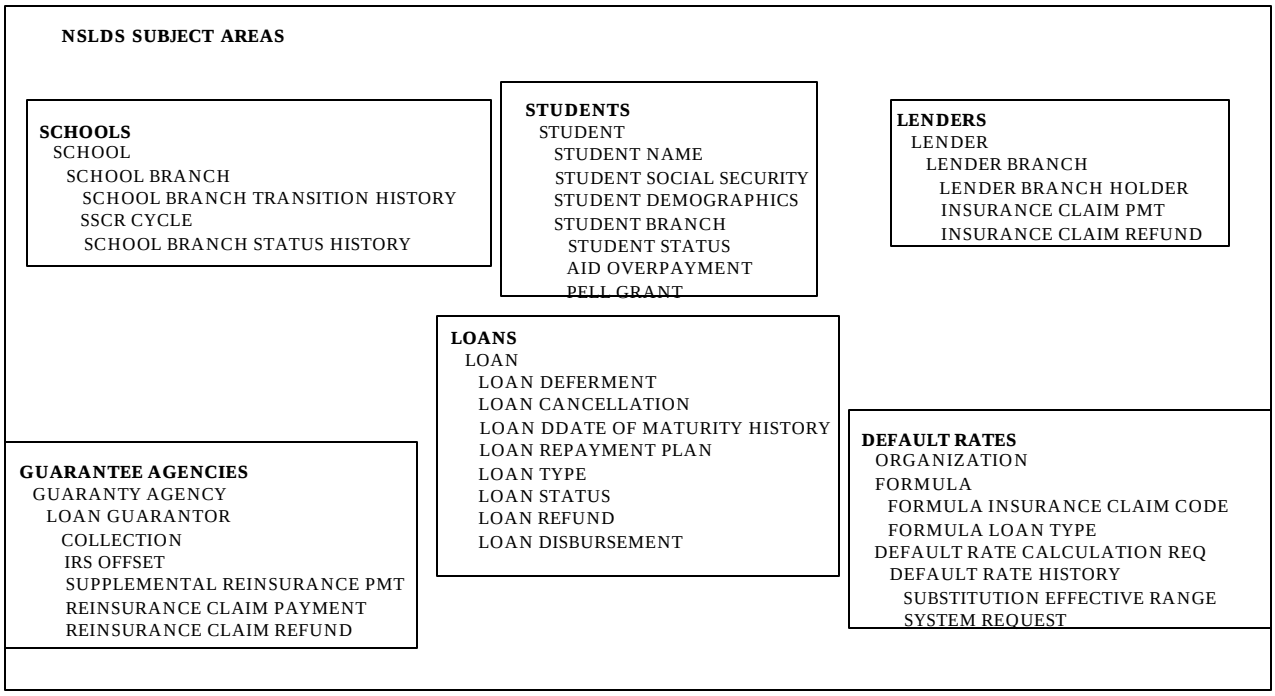
Online Statistical Abstract (STAB) Database – The STAB database was created in response to research teams asking for a sample of NSLDS data that is random and statistically valid. STAB is created (refreshed) quarterly and consists of a subset of data representing a cross-section of both Active and Archive databases, and by volume is less than 5% of the total data set.

End-user Database – This database allows users to store results of queries so that it is possible for users to view, analyze or share the data at their convenience. Users use the query tool QMF (Query Management Facility) to develop queries.

The Active, Archive and the STAB databases share the same database design, but with different data content. The End-user database emphasizes table space, rather than design.

Subject Model

Shown below is a high-level subject model of how NSLDS' database is organized. There are 7 subject areas that comprise NSLDS: Loan, Default rate, Student, School, Guaranty Agency, Lender, and Support. The Loan and Default rate subjects are central to the model. The Support subject area contains cross-reference, lookup, and work tables. See Appendix F (Conceptual Entity Model by Subject Area) for a more detailed review of subject areas. See Appendix A (Attributes by Business Function) for more detail at the attribute level.



The remainder of this section describes how the COOL:Gen tool and DB2 together play an active part in database design, creation and optimization.

Database Creation

Virtually all design and development of NSLDS was completed using COOL:Gen. COOL:Gen is a CASE tool that tightly integrates database design and creation with application design and code generation.

The tool was central in creating NSLDS database, and continues to be used in maintaining both the database and application. COOL:Gen works with the database server to create the physical data structures. In this case, the database server is DB2. COOL:Gen generates the data

dictionary upon completion of the physical data model, which serves as the authoritative database documentation.

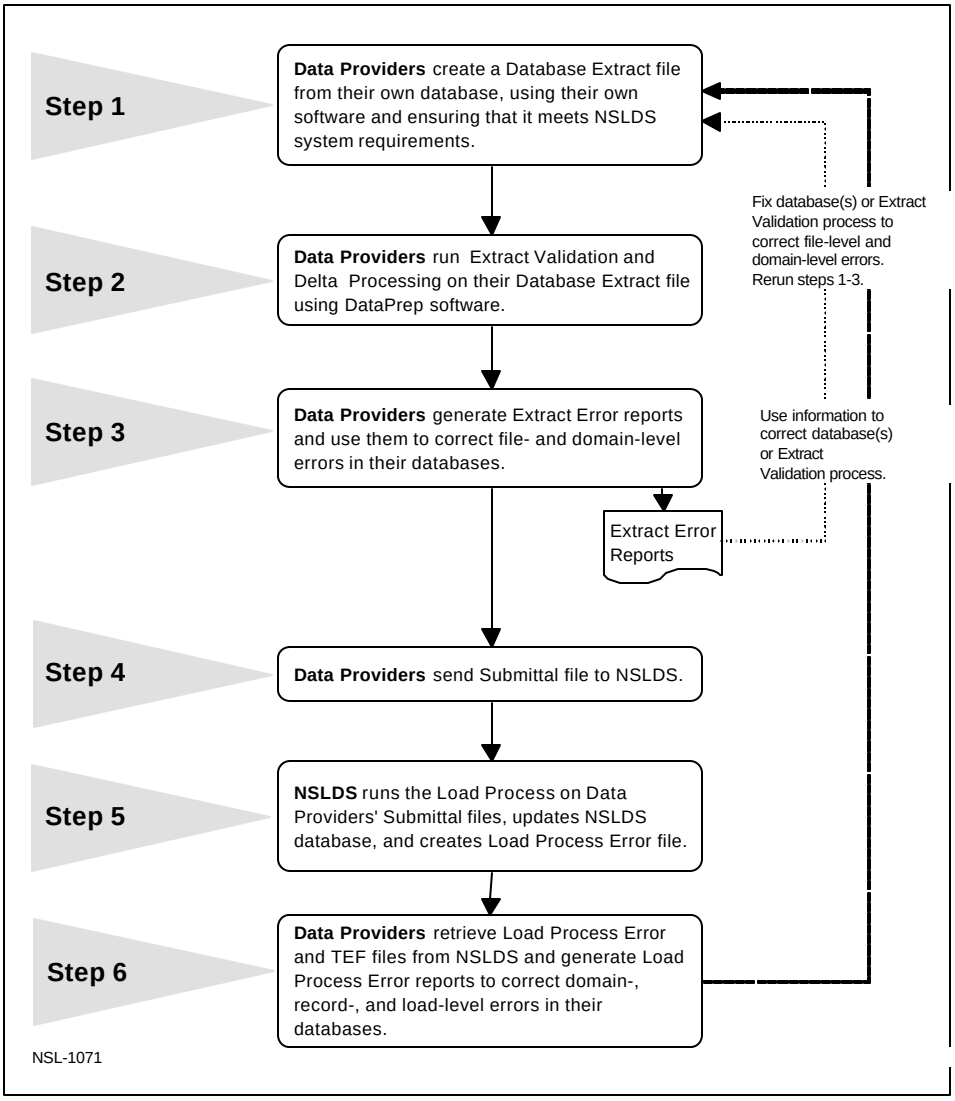
Database Optimization

The NSLDS database was optimized using either DB2 tools and techniques or third party tools such as Platinum Plan Analyzer or INSIGHT. The DB2 Optimizer and the Platinum Plan Analyzer tools aided in determining query access paths to tune table indexes. INSIGHT was used to monitor both the database and application code performance. Additionally, optimization techniques, such as type of tablespace design (simple, segmented, partitioned) and disk space allocation were used in tuning NSLDS' performance.

Data Population Architecture

A brief overview of the data population process: NSLDS receives data from a large number of external sources and five internal SFA sources, which are described later in this section. Since external systems vary tremendously in format and architecture, NSLDS designers asked all sources to perform their own data extracts and transformations, and send the information to NSLDS in prescribed file format. NSLDS designers, however, provided transformation software to aid external sources to provide clean data.

When data files come in, NSLDS re-runs the transform logic to double-check the data quality before attempting database updates through the load process. In other words, the extract and transform logic is run by data provides at their sites, with the output format that NSLDS expects, and the tranform and load logic is run by NSLDS.



The remainder of this section discusses:

- data sources, what they provide, and how often
- extract, tranform and load processes
- data quality assurance
- scheduling

There are a number of appendix sections with supporting detail which are referenced where appropriate.

Data Sources

NSLDS obtains its data from schools, guaranty agencies and SFA systems, including the:

- Direct Loan Servicing System
- Debt Management Collection System
- Pell Grant System
- Postsecondary Education Participant System
- Central Processing System

See Appendix I (NSLDS Sources and Uses) for a source diagram.

Below is a quick description of what information is gathered from the different data sources:

FFELP Loans Data – guaranty agencies or their servicers collect FFELP loan data held by either their lenders or themselves, and send that information to NSLDS.

Perkins, Pell, and FSEOG Data – school or their servicers send information on Perkins loans to NSLDS. However, Pell, FSEOG and Perkins overpayments can also be entered through NSLDS' online interface.

Student Enrollment Data – a student enrollment status roster is updated in full at least once a semester but can be scheduled more frequently by a school. Student enrollment changes between submissions are entered online within 30 days of the student's enrollment change.

Pell Grant Data – RFMS sends NSLDS Pell grant update information.

Debt Collection Data – Federal Family Education Loan / Debt Collection System (FFEL / DCS) sends NSLDS data on all loans held by SFA in the Debt Collection System portfolio, including FISLs, Perkins loans, FDLP loans, and FFELP loans. FFEL / DCS also supplies the GA code table and lender code table information to NSLDS.

Direct Loan Data - Direct Loan Servicing System (DLSS) sends NSLDS direct loan information from the first disbursement until they are repaid or the borrower becomes disabled, enters bankruptcy, defaults, or dies.

Basic School Data – Postsecondary Education Participants System (PEPS) sends NSLDS updates to the school table. NSLDS also provides a means to convert Pell Institution codes, FFEL School IDs, and Campus-Based School codes to OPE IDs.

The table below shows the sources, the frequency of updates to NSLDS, and loan, grant or record volumes.

System	*Submittal Frequency*	*Largest Submittal Volume*
Guaranty Agencies	Monthly	> 17,000,000 loans (USAF)

Perkins Schools	Monthly	> 1,000,000 (AFSA)
CPS Demographics	Was Twice a Year, Now Quarterly	> 10,800,000 student records
RFMS Pell Grants	Daily	~ 750,000 grants
PGRFMS Pell Grants	Monthly	~ 500 grants
Debt Collection System	Monthly	> 4,500,000 loans / grants
Direct Loan Servicing System	Monthly, soon to be Weekly	> 12,500,000 loans
PEPS	Weekly	> 332,000 records

Functionally, NSLDS captures and stores almost all information related to Title IV student aid information. The following is a sample of the information stored in NSLDS:

- student / borrower identifiers - social security number, date of birth and name
- loan information - entire life cycle of a loan, from origination through final payment, cancellation, discharge or other final disposition, loan amounts, educational status, disbursements, balances, loan status, collections, claims, deferments, refunds and cancellations
- student enrollment information - including school(s) attended, anticipated completion date, enrollment status and effective dates
- student demographics - course of study, dependency, citizenship, gender, family income, expected family contribution, and address
- grants and Perkins information - Pell Grant amounts and dates, Federal Supplemental Educational Opportunity Grant, and Federal Perkins Loan Program overpayments
- guarantor, lender, servicer and school demographics - profile on guaranty agencies, educational institutions, financial institutions, loan servicers, and third party data processing servicers
- user profiles and contact information for student aid professionals at the Department of Education, guaranty agencies, educational institutions, financial institutions, loan servicers, and third party data processing servicers
- NSLDS also maintains historical information for all functional area.

Design Considerations

Population of student aid information from literally thousands of independent sources is a fundamental requirement for NSLDS. Achieving this goal requires several important considerations:

- The type of loan is unique to a data provider type; for example, a GA would never be responsible for a loan record for a Perkins Loan, or a school would never be responsible for a loan record for a PLUS Loan.

- The regulations are unique for each loan or grant program, therefore the data elements and validations are unique for each aid program.

- The data collected from external data providers is dependent on the business functions that will use the data. The data elements were reviewed and approved by OMB on the basis of the business function that would be served and that the benefit justified the reporting burden.

- The identifiers chosen for a loan or grant, must be data elements that a majority of data providers collect and store in their database, plus data elements that add uniqueness to a given occurrence. This is because there is no national identifier for loans or grants and data providers do not use the same data elements to uniquely identify a loan or grant.

- Loans move from system to system. Lenders buy and sell FFEL loans everyday on the open market. Guaranty agencies can close causing the loans they hold to be transferred to another guaranty agency. Many schools turn over their Perkins loans to loan servicing companies when a student enters repayment. Schools and guaranty agencies can outsource their data processing to service bureaus and change servicers at will.

- Cumulative amounts are reported, ensuring that the affect of all events and changes during the interval since the last valid update are reported regardless of the reporting frequency. It also ensures that both current and past period amount corrections posted to the data provider's system are automatically captured as a change and sent to NSLDS.

- The data must reflect what is in the data provider's database. All corrections must be made through the data providers system and first incorporated in his database before NSLDS is updated. NSLDS must be an accurate reflection of the data provider's data.

- The NSLDS data must be left in a consistent state after the update. Partial events or inconsistent events cannot be accepted since users would never know which events were accepted or rejected.

- Data Providers must be able to roll back or undo events that are reported but subsequently deleted in their system. For example if a loan is placed into repayment status but the lender later discovers the student was eligible for a deferment or forbearance.

- Following these and many other high level design concepts a Data Provider Instructions (DPI) document, has been prepared for each source system at a very detailed level that provides a rule book and guide for extracting, transforming, formatting, editing and transmitting data.

Extract, Trasform and Load Processes

NSLDS has a process in place for data providers to send their data to SFA. This process adresses the two basic areas of assuring data quality and getting regularly scheduled updates.

Data providers are asked follow these guidelines when they extract and send data to NSLDS:

- send new loans and updates to existing loans per the frequency defined by NSLDS
- send all loans in the portfolio from the time a loan is guaranteed until either the loan is closed and successfully reported to NSLDS or the loan is transferred to another organization or the loan is subrogated to SFA
- do not send closed loans already reported and successfully loaded into the NSLDS database, based on the report NSLDS creates
- the data file must contain a header record and a detail record for each of the loans in the portfolio; the file may contain Past Period Change records if historical data in NSLDS is being modified; it should not contain a Trailer record, since the software will generate a Trailer record for the Submittal file
- once a data file has been created do not alter the data in it, or in any of the files subsequently created from it
- each record in the data file must go through the NSLDS-provided Extract Validation software; errors identified during this process must be corrected in the data provider's databases, not in the Database Extract file or current Submittal file
- loan information must be supplied for all the fields included in the Data Provider Instructions, unless the associated event has never occurred
- detail records in the data file must be at the loan level
- the data must be not be earlier than 14 days before the scheduled load date
- many of the fields identified by NSLDS must be meet field content rules such as valid status, conditional calculations, etc.

NSLDS has provided data providers with validation software that run on either the mainframe computers running under the OS/390 operating system with the LE environment or the PC. If the data provider does not have either of these platforms, they are expected to develop their own software based on NSLDS specifications.

The data provider runs through the steps of extracting data, running the validation programs, running error reports, creating deltas (GAs only), and transmitting the submittal file to NSLDS.

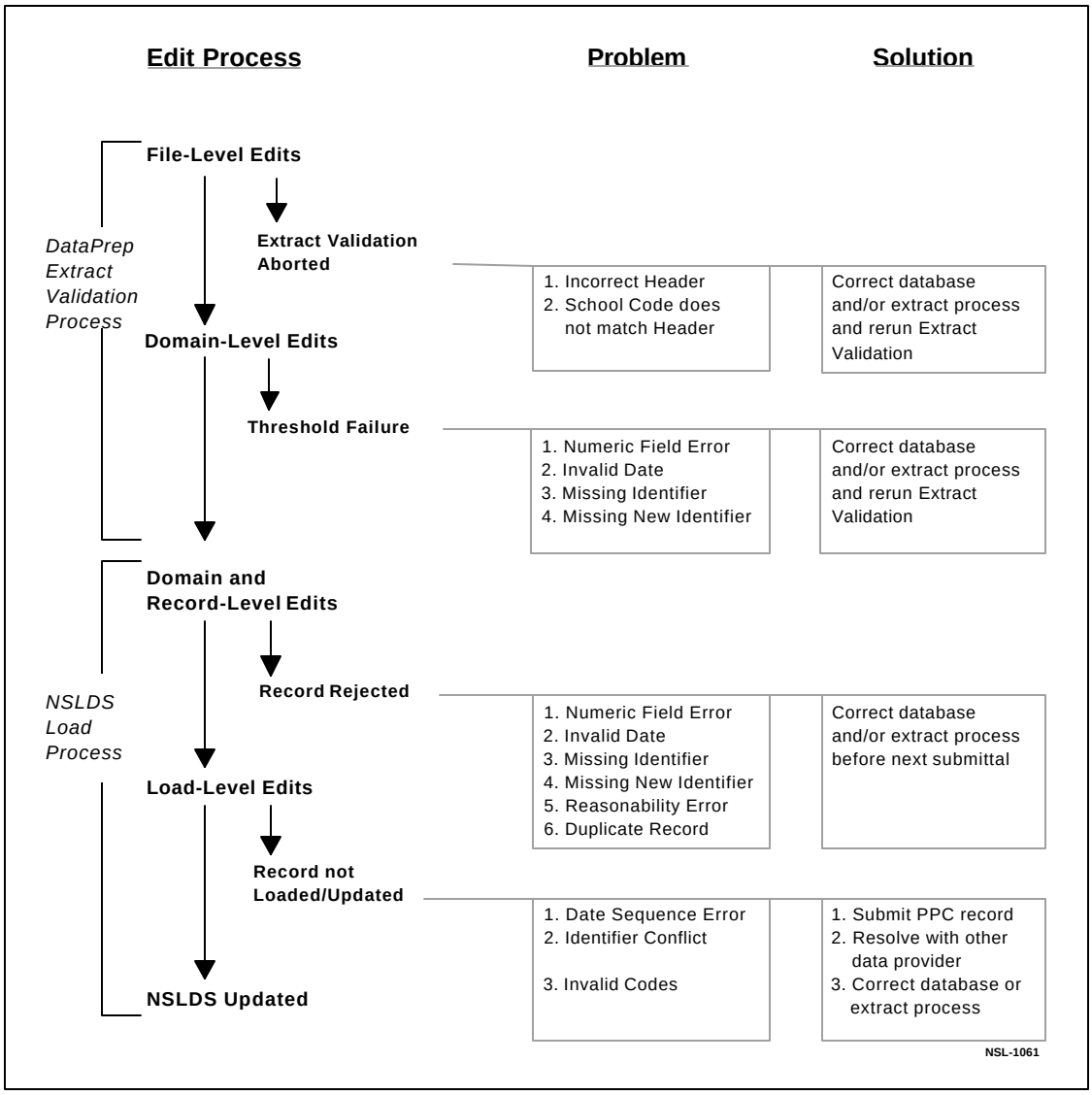
NSLDS runs throug the steps of pre-processing the submittal file (receipt and staging), load process verification and update process, creating and sending error files, and transmitting the error files back to the data provider.

Data Validations

Data validation is considered part of the transformation processing. NSLDS uses a three-step validation process to assure the quality of the data that enters the repository:

> ***Level 1*** – file and field (domain) level validations are run initially at the data provider site and then again at NSLDS. This process ensures that the data file is complete and in tact. Field level validation assures that field contents are of the expected type (eg. date in a Date field).
>
> ***Level 2*** – delta processing is run only at guaranty agencies due to their large monthly volumes. The delta processing software is provided to guaranty agencies by NSLDS to reduce the volume by extracting only changed records.
>
> ***Level 3*** – record and load level validations are run only by NSLDS to ensure the validity of record content in record-level validation, and database content and referential integrity at the load-level validation. Record and load level validations include table lookups, calculations, restatements and reconciliations of record or field content before it enters NSLDS.

The diagram below shows validation processing.



File Formats

NSLDS has defined 7 file formats for incoming data from schools, guaranty agencies and internal SFA systems. All schools use a format, guaranty agencies use a different format, and each of the 5 SFA systems have a format each. However, all data files have the three components: a header record for basic information about the file, detail records for adds / updates / deletes, and footer record for marking the end of file.
Additionally, detail records are marked to indicate whether it is a correction record or new with "past period change" mark.

Scheduling

NSLDS has arranged submittal dates with the various data providers to accommodate the update schedule to their operational systems, the availability of processing windows on their systems, and to balance the daily scheduled loads to the NSLDS system. A program is run each year, usually in November, to compute and distribute the submittal schedules for each data provider for the upcoming year. The schedule can also be viewed at any time by selecting the Data Provider Schedule link on the Organization page of the NSLDS web site *www.nsldsfap.ed.gov*. See appendix G (NSLDS Production Schedule 2000) for an example of an annual NSLDS production calander.

A data provider's Submittal file should arrive at NSLDS no later than 1 p.m. central time the day before it is scheduled for processing. The data it contains cannot have been extracted from the data provider's operational database more than 14 days before the scheduled load date. The data provider's submittal must be received within a window that runs from 14 days before the Submittal file is scheduled for processing to 13 days after. If a Submittal file is not received by 1 p.m. Central Time the day before the scheduled load date, NSLDS contacts the data provider to remind them their submittal is due.

If a Submittal file is not received within the data provider's submittal window, NSLDS contacts the data provider again to inform them that their Submittal file cannot be processed that month (skipped submittal) and that they must send a new Submittal file the following month.

NSLDS receives Submittal files continuously during the day. As tapes, cartridges, or electronic Title IV WAN files are received the files are validated and the data is loaded to disk for later update processing.

The update processing job itself takes advantage of multi-thread design and will run 12 times in parallel allowing processing of approximately 7 million loans per CPU hour. See Appendix E (Submittal Processing Volumes) for sample load statistics from a recent 10 day period .

Application Architecture

Fundamentally, NSLDS has two application areas: system-to-system, and system-to-user. System-to-system is when NSLDS sends data to other systems. System-to-user is when NSLDS presents its data to end-users through views, reports or queries.

System-to-System Application Architecture

Central Processing System (CPS)—CPS sends NSLDS a daily feed identifying applicants for new or additional Title IV aid. NSLDS returns to CPS financial aid history data for any applicant found in the NSLDS database. From this data, CPS generates an Expected Family Contribution and a Pell Grant Index. CPS sends NSLDS student demographic data from the FAFSA application twice a year. See appendix G (NSLDS Business Functions, Hi-level Processes, and Facts by Dimension) and reference the first section on Aid Eligibility.

Post-secondary Educational Processing System (PEPS)— The weekly interface between the Postsecondary Education Participants System (PEPS) and NSLDS supplies the information necessary to support population and maintenance of school-related data such as new schools, mergers, splits, closures and identifying information such as OPE code, name and address. Based on this school data, twice a year (drafts in the Spring and Officials in the Fall), NSLDS calculates default rates and returns the results to PEPS which ultimately decides school eligibility. See appendix G (NSLDS Business Functions, Hi-level Processes, and Facts by Dimension) and reference the fourth section on Default Rates (School Eligibility).

Schools/GAs/Lenders— The NSLDS forwards a SSCR Roster file (Student Status Confirmation Report) to a school's Title IV mailbox based upon the SSCR cycle previously established by the school. If the school does establish a reporting cycle, the default setting is bimonthly. The school may identify a servicer to process their enrollment information. If this occurs, the SSCR roster is sent to the servicer destination. The school, or its servicer, is required to update the status of the records contained on the roster within 30 days of receipt. The NSLDS sends this updated enrollment information to GAs, and to the FDLP servicer. GAs supply this information to lenders and lender servicers. See appendix G (NSLDS Business Functions, Hi-level Processes, and Facts by Dimension) and reference the second section on Enrollment tracking.

System-to-User Application Architecture

The remaining functions (Student Aid Tracking, Financial Management, Budget Forecasting and Credit Reform Act Support, Research, Policy Analysis and Performance Assessment) are delivered through various paths using on-line and batch techniques including web display, pre-defined reports and ad-hoc querying capability. Summary and detail student financial aid history is available at *www.nsldsfap.ed.gov*.

Approximately 50 pre-defined reports used by Schools, GAs and ED are available through on-line web request and are generated nightly (or sooner) for delivery through TIVWAN. The pre-defined report selection web page has a dynamic design (unlimited selection of report

parameters and sort selection criteria) allowing new report set-up without any programmer involvement. See appendix C (NSLDS Pre-Defined Reports) for a listing.

For end user ad hoc queries, sophisticated and detailed DB2 "View" definitions are defined for all physical NSLDS tables. These "views" were designed to simplify user access and to allow efficient access to NSLDS tables (prevents unnecessary joins to high volume tables). IBM tools such as QMF / SQL are available for end-user access as well as PRF (Platinum Reporting Facility). Raytheon also maintains an "Info Center" group to assist users with difficult queries.

Application Environments

From initial development in the COOL:Gen Development model, to the User Acceptance model and finally to the production model, all objects are controlled by COOL:Gen Business objects (Data and processes) are initially defined in the development model and then migrated (as single objects or as aggregate sets) to User Acceptance and Production. Code and Database objects (i.e. DB2 Data Definition Language or DDL) cannot be generated until the logical or business objects have been migrated.

COOL:Gen provides for tight integration of all objects (including physical data objects) between all groups (application development, Model Management, Configuration Management, Database Administration, etc.). For example, an entity defined in the logical data modeling tool is transformed into a Technical Design "Record" which will become a DB2 table with no loss in logical or physical properties. All groups have one single definition for an object regardless of the model or phase of development

This tight integration dictates that no business or design considerations will be lost. The use of aggregate set migration also reduces the opportunity for error and allows for smooth transitions between environments. The tool provides a sophisticated consistency checking feature (using artificial intelligence concepts) to make sure that no Information Engineering Methodology rules have been violated. The consistency checking feature can be executed at all phases (Analysis, Design, Construction).

NSLDS has a version control strategy that allows simultaneous development of different versions if necessary. The majority of NSLDS software is executed on a central IBM mainframe so there is no real need to track and maintain different versions except for the extract software (windows 95, 98, NT, IBM) provided to Schools and GAs (NSLDS "DataPrep" software) which is maintained by Configuration Management

See appendix J (NSLDS Environment Review) for a diagram of how the current NSLDS environments (development, test and production) are maintained including the overall version control strategy. The diagram also demonstrates the chronological history of NSLDS environments by discussing several short term environments (SOT and System Test).

The application software is written using the COOL:Gen action diagramming language with COBOL selected as the target language on the mainframe and "C" selected for the PC. The NSLDS web interface uses Netscape or Microsoft's Internet Explorer as the browser (including

ASP) linked to a web server in Meriden, CT using Microsoft's IIS (Internet Information Server) software and COOL:Gen middleware (COM:Proxy and Communication Bridge). COOL:Gen has driver diagrams (procedures) and called diagrams (action blocks) and external action blocks (EABs) which are hand-written routines (COBOL, C) to allow access to non-COOL:Gen accessible objects (flat files, IMS files, etc.). NSLDS has no DB2 stored procedures. RI triggers are enforced through COOL:Gen, DB2 triggers are not used.
See appendix J for statistics on NSLDS application software volumetrics.

The infrastructure groups (Configuration management, model management, Data Base Administration, QA/QC, etc.) require a minimum of two weeks to review changes before migration to production. This is done through a Test Readiness Review (TRR) and a Production Readiness Review (PRR). Bypassing TRR or PRR requires SFA approval.

Security
NSLDS uses the Role Based Access Control (RBAC) concept as described in the Modernization Blueprint for determining access to specific data. Access is based on an individuals job function, location / organization, and can be restricted to only certain action commands.

Security is controlled through the issuance of user logon IDs and passwords. IBM's Resource Access Control Facility (RACF) is the primary security product. Database access and query tool use are controlled by DB2 authorization and views.

Operations Architecture

Application Maintenance and Enhancement

NSLDS is maintained using the COOL:Gen CASE tool. Changes and enhancements are assigned to a future planned version and release. Each future version / release has an associated COOL:Gen model in the COOL:Gen Central Encyclopedia (CE) that has been populated with the necessary COOL:Gen objects necessary to accomplish all of the changes packaged into a particular version / release. The CE automatically controls the checkout of COOL:Gen objects to prevent one developer from overlaying another developer's work. Thus multiple releases can be in-work concurrently without the usual manual coordination. The impact of changes to logical or physical database structures is assessed using the COOL:Gen CASE tool as well as BMC Change Manager.

The design diagrams are modified by the developer and debugged using the COOL:Gen Test-from-Design features. When the developer is satisfied with the initial operation, the target computer type, operating system, database, and source code language is selected. Source code is automatically generated and optimized for the target environment, compiled for efficiency, and the load module libraries updated in the Development environment. Further testing is performed by the developer and any changes are made to the COOL:Gen diagrams, not to the source code. The computer generated source code is never used for maintenance or changes.

When the developer is satisfied that the application is debugged the COOL:Gen model objects are migrated to the Test environment and the source code regenerated, compiled, and the load modules installed. The Testing group tests the application using prepared test scripts. Any defects are corrected by the migration of the affected object back into the Development model and environment for correction by the original developer. When the application passes unit testing and the integration testing of all of the changes for the new release are working, the load modules are copied to the User Acceptance model and environment. The load modules are copied rather than re-generated by COOL:Gen to ensure that the application as tested is exactly the same as the application presented for User Acceptance. When the user accepts the application, the load modules are copied to the Production environment and the Production model. The old production model is backed up before the update to permit an immediate rollback if problems occur in the new release. A Fix model is maintained in sync with the Production model to facilitate temporary emergency changes. If changes are made to the Fix environment then the permanent changes must be made to the Development model as soon as possible and migrated through the normal test and acceptance procedures into Production.

The Pre-screening and Post-screening Aid Eligibility functions and Student Aid History functions are regularly revised every year because of changes in eligibility rules and changes in the FAFSA aid application form or the resulting ISIR report's content. The Default Rates function is revised each year because of changes in the formulas, exclusion rules, or changes in the loan programs to be included. The Monitoring of GA and Lender Billings for Reasonability function changes whenever changes are made to the ED Form 799, 1130, and 1189 or the rules for payments are revised. New loan status codes, deferment types, insurance and re-insurance

claim reasons, and other NSLDS data can be impacted by changes by SFA rules, legislation, or the ED Policy group. Changes to the data elements can have wide ranging impact including changes to the Data Population software, changes to edits, changes to the database updates and changes to functions that use the data element. Usually changes can be anticipated or foreseen well in advance and incorporated into a future release for orderly implementation.

Disaster Recovery Preparations

NSLDS must demonstrate a complete Disaster Recovery plan each year at an offsite data center selected by ED. The Disaster Recovery demonstration begins with an equipped shell data center and does a full restore of the NSLDS operating system, database, and application software. Communications recovery usually only covers critical network functions as decided in advance by SFA. The time required to fully restore NSLDS operations is dependent on the CPU and the number of tape drives available, but in past Disaster Recovery exercises with equipment equivalent to the production site, the normal time has been from 4 to 8 hours. NSLDS has never required a real disaster recovery of the production system, only the annual demonstrations.

To permit emergency recovery, any changes to the NSLDS database are logged to permit full recovery to currency at any time using DB2 journaling and recovery processes. The high volume update jobs use CDB Restart+ to facilitate the automatic recovery of an update job without starting over if a failure occurs in the middle of a job. Image copies of the DB2 database are performed after the Monday - Friday night data provider updates while the database is quiesced. Image copies are run in parallel against each database partition and are limited by the number of available tape drives. The Image copies usually require about 5 hours to complete. Full disk pack backups are performed every Sunday night and require about 8 hours to complete, but can be run in parallel and the time depends on the number of tape drives available.

NSLDS has a message and notification feature on its web sites and logon screens, which can be used to alert users to upcoming scheduled activities. This feature could be used to notify users of planned outages. In addition, in the event of a system outage, ED, the VDC help desk, and the NSLDS Customer Service Center are notified so that the help desk representatives know the system status and can assist users that may call.

The NSLDS design provides for an Archive database to provide online access to closed student account data while reducing the size of the Active database. ED accepted the design, but archiving is not been performed at the direction of SFA. The historical data required to support the NSLDS business functions has changed over the years as ED rules and regulations have changed. The implementation of archiving would require a review of the current user needs and legislative requirements to determine if the original archiving rules still apply.

Service Levels

The Service Level Agreement for NSLDS is between SFA and the Virtual Data Center contractor CSC and we did not receive a copy of the SLA document in time to include it in this deliverable. The service levels that are known are:

- NSLDS operates 24 hours a day, 7 days a week, 365 days a year.
- Customer Service support is available from 7:00 AM to 7:00 PM Eastern time, Monday through Friday except for Federal holidays.
- The system is available at all times except during the weekly scheduled IPL of the system at 4:00 AM Monday, which lasts about 20 minutes.
- There is no fixed batch window, but data provider updates and database backups are scheduled during the off-hours between 6:00 PM and 8:00 AM. The database updates and user access can occur concurrently and occasionally the updates will be started early or run late if an extremely large volume exists.
- Routine database maintenance, re-organizations, performance tuning, adding disk space, and adding or removing indexes is performed during a weekly scheduled DBA maintenance window each Saturday night. Action is only taken if needed, but an 8 hour window is reserved in case it is needed.
- Hardware and system software maintenance is performed during a weekly scheduled VDC maintenance window from midnight to 4:00 AM each Monday morning. Action is only taken if needed, but the 4 hour window is reserved in case it is needed. As mentioned above at a minimum a system IPL is performed. Occasionally if a major upgrade is planned, the VDC will request additional maintenance time after midnight Sunday.

VDC Hardware and System Software Maintenance

The hardware and system software are a shared resource at the ED Virtual Data Center and are maintained by Computer Sciences Corporation (CSC). Changes to the hardware and system software are discussed with the application development team from Raytheon and the impact on the users or developers assessed. The Configuration Managers from both the CSC VDC team and the Raytheon development team coordinate and control all activities that affect the application software, systems software, or hardware in the Development, Test, User Acceptance, Training, and Production environments.

Analysis

NSLDS was originally designed to collect and centralize Title IV student aid data from both external and internal sources. Since then, there has been increasing levels of demand for access to that centralized information. Many of these demands have been met by adding to the original design and processes, or by workaround methods.

Fundamentally, NSLDS contains the same components as a classic data warehouse, in that it has an extract-transform-load component, a database component and a query and reporting component. However, the database design is intended for online transaction processing (OLTP) instead of analytic processing. Based on this recognition, we have made a few suggestions that are intended as ideas to improve NSLDS' architecture.

ETL Process Re-engineering - NSLDS receives data files everyday, since there are hundreds of sources. Today, guaranty agencies are the only data providers who run delta processing, sending only records that have changed from the last submittal. There are two process improvement opportunities here:

- All sources should either send deltas to NSLDS or NSLDS must have a way of recognizing deltas before the ETL process occurs. For example old data files can be compared with new incoming ones before going through the ETL process.

- We should look for opportunities to simplify the scheduling of data files. Currently NSLDS has to make updates successfully every night. Identify areas where we can group like files for ETL and updates.

Potentially Redesign Data Model – NSLDS' update process adds, changes or deletes fields in the database. This is costly in terms of performance of the update process. In the data warehousing paradigm, the update process always adds with a new date/time stamp. This serves to maintain history while reducing the burden on the ETL processing. The downside is it takes more disk space.

We should also look for opportunities to redesign the model for better information delivery to the users. NSLDS' database is organized to serve OLTP functions, but there are no OLTP functions it serves. The database should be organized as a "star" or a "snow flake" schema to better deliver the information to both systems and users.

Use Archiving – Correct use of archiving will make the database size smaller and more manageable to the ETL process and the information delivery process. Today NSLDS does not archive for the reason that research and other communities need the full data set online.

Pre-run Queries – After the ETL process, pre-running frequently used queries from systems and users will deliver information faster. NSLDS, for example, waits for CPS to send student ID

information before it runs student eligibility queries. These queries can be pre-run, if the ETL process scheduling and the data set is managed properly.

Use COTS Tools – COTS tools such as Informatica for data population and MicroStrategy for querying and reporting can greatly simply application maintenance. Today COOL:Gen is at the center of NSLDS, inter-twining the data modeling, the database and the application.

De-couple the Data from the Application - SFA wants to move toward de-coupling the data from the application. Separating the data from the application logic will enable SFA to select best-of-breed technical components and use them in a "plug-and-play" fashion without affecting the whole system.

Use Open Technologies and Tools – The new enterprise standard at SFA is to make consistent use of open technologies and tools. Most technology providers adhere to these standards in the areas of web, database, electronic data exchange and access to data. Use of open technologies is important from a skills availability standpoint also.